# Benchmarking Intelligent Service Robots through Scientific Competitions:The RoboCup@Home Approach

**Dirk Holz**
Autonomous Intelligent Systems Group,
Department of Computer Science VI,
University of Bonn, Germany,
dirk.holz@ieee.org

**Luca Iocchi**
Dept. of Computer, Control, and
Management Engineering
Sapienza University of Rome, Italy
iocchi@dis.uniroma1.it

**Tijn van der Zant**
Dept. of Artificial Intelligence
Faculty of Natural Sciences
University of Groningen, the Netherlands
tijn@ieee.org

## Abstract

The dynamical and uncertain environments of domestic service robots, which include humans, require rethinking of the benchmarking principles for testing these robots. Since 2006 RoboCup@Home has used statistical procedures to track and steer the progress of domestic service robots. This paper explains the procedures and shows outcomes of these international benchmarking efforts. Although aspects such as shopping in a supermarket receive a fair amount of attention in the robotics community, the authors believe that a recently implemented test is the most important outcome of RoboCup@Home, namely the benchmarking of robot cognition.

## Introduction

Benchmarking intelligent robotic systems is a fundamental task necessary to successfully deploy intelligent robots in everyday activities. Moreover, benchmarking is of utter importance for the marketability of intelligent robots.

In this paper we focus on domestic and service robots (DSRs). They assume a large variety of forms and intend to solve several types of problems. Moreover, DSRs have a primary objective of fluid interaction with people in typical human environments (e.g., a house or an office). The development of intelligent DSRs requires the integration of many capabilities coming from various research fields: robotics, artificial intelligence, human-robot interaction. Two issues make the development, and consequently benchmarking and performance evaluation, very difficult. These issues are 1) the interaction with humans in real environments and 2) the integration of several capabilities coming from different research fields.

Developing methods to evaluate integrative research and realize good benchmarks are thus more difficult in a domestic and service robot application as: 1) humans and real environments must be explicitly considered in the benchmarking methodology and 2) there is large assortment of tasks that can be accomplished by a DSR.

In this paper we present a benchmarking methodology for domestic and service robots that addresses the previously mentioned problems. In particular, this methodology

is based on scientific competitions and has been adopted in the past years by the RoboCup@Home competitions. It allows for testing robots on many integrated tasks (as opposed to single functionalities) in real environments as well as testing interaction with humans (non-developers of the tested system).

More specifically, this paper presents the "benchmarking through scientific competitions" approach of RoboCup@Home, showing the actual method for evaluating integrative research, the results of this evaluation over the years, a specific benchmark for the reasoning capabilities of the robots and a discussion about how teams solved and integrated the many problems that must be handled for a successful implementation.

## RoboCup@Home methodology

RoboCup@Home[1] (Wisspeintner et al. 2009) is the largest and most common benchmarking effort of integrated robotic systems in the fields of domestic service robotics and human-robot interaction.

RoboCup@Home is a competition that usually runs in a 6 days period (including 1.5 days of setup) in a test arena that is configured as a realistic representation of a home environment. However, some tests are also executed in the competition venue or in real environments nearby, such as restaurants, shopping malls, etc. The competition scenario (i.e., both the test arena and the real environments) is not specified beforehand. For the competition scenario, teams have a short setup time (about 1 day and a half) before the competition starts to prepare for the particular scenario and to perform calibration, etc. For the real environments (real shops in 2010 and 2011, and a restaurant in 2012), there is no setup. Here the robots must properly function immediately in a completely unknown environment.

The competition includes several so-called *tests* that test for particular capabilities and their successful integration. In addition, the competition features open demonstrations where teams can showcase their latest research results or draw the attention to problems that have not yet been considered in RoboCup@Home, possibly leading to future tests for the competition.

---

[1]http://www.robocupathome.org/

In the regular tests many different functionalities such as navigation, localization, object recognition, manipulation, and speech and gesture recognition are simultaneously tested in an integrated manner. Good results are obtained only when a complete system, implementing all the desired abilities, is successfully demonstrated during the competition. By requiring teams to participate in many different tests, we force the development of general solutions as hacking specific solutions for many different problems would be ineffective.

In order to avoid static benchmarking, where the risk of progressing towards local optimum solutions may arise, tests are changed on a yearly basis. Tests change according to the decisions of the experts in the Technical Committee that integrates both the suggestions by the teams and the results of the statistical analysis described below.

The main ingredients of the RoboCup@Home methodology can be summarized as: 1) benchmarking integrated systems (rather than single functionalities); 2) performing several different tests; 3) changing tests over time. In our opinion, these features, which are unique in the field of benchmarking intelligent service robots, are key factors able to drive researchers towards developing high quality and robust scientific solutions, while penalizing specific ad-hoc implementations.

However, the previously described features also introduce problems that must be considered: 1) benchmarking integrated systems does not allow an easy evaluation of the performance of the single components of a complex system; 2) dynamicity of the benchmark makes it difficult to relate the temporal evolution of performance (Feil-Seifer, Skinner, and Mataric 2007).

These problems are even more acute regarding tests that require the robot to operate in the real world.

The answer to these problems provided by the RoboCup@Home methodology is to perform statistical analysis on the results. This is possible due to the participation of many teams in many tests, thus providing a good data for analysis.

In the remainder of this section we will first describe the implementation of the RoboCup@Home competitions and then discuss the results.

## Implementation of the competition

The competition is organized in several tests that must be accomplished by the robots. These tests are integrated tests, thus each test comprises a set of functionalities that must be properly integrated to achieve good performance. However, the scoring system allows for giving partial credit if only a part of the test is achieved. More specifically, the total score of a test is the sum of certain partial scores. Each partial score is related to a specific functional ability (see next section).

Another important issue of this organization is that different tests often require the use of the same functionality but in different conditions. This enforces the development of robust and flexible solutions as a specific ad-hoc solution for one test may not be appropriate for other tests, and developing as many ad-hoc specific solutions as the number of tests is not feasible.

Finally, tests change over the years in order to avoid development of solutions that converge to "local optimum" performance, i.e. solutions that are overspecialized and optimized for a single test or scenario.

## Statistical analysis

One important feature of the RoboCup@Home methodology for benchmarking analysis (see (Wisspeintner et al. 2009) for details) is the measure of performance advances over the time of an evolving benchmark. This is obtained by analyzing the partial scores of each test, which are associated to specific functionalities, and by evaluating the amount of score gained by the best teams for each functionality in each test. In this way we can measure the average increase of performance of the given skills over years, even when changing the tests.

The functional abilities that have been considered in the competitions are the following:

- *Navigation*, the ability of path-planning and safely navigating to a specific target position in the environment, avoiding (dynamic) obstacles.

- *Mapping*, the ability of autonomously building a representation of a partially known or unknown environment on-line.

- *Person Recognition*, the ability of detecting and recognizing a person.

- *Person Tracking*, the ability of tracking the position of a person over time.

- *Object Recognition*, the ability of detecting and recognizing (known or unknown) objects in the environment.

- *Object Manipulation*, the ability of grasping, moving or placing an object.

- *Speech Recognition*, the ability of recognizing and interpreting spoken user commands (speaker dependent and speaker independent).

- *Gesture Recognition*, the ability of recognizing and interpreting human gestures.

- *Cognition*, the ability of understanding and reasoning about the world.

As already mentioned, these functionalities are distributed throughout the tests and associated to partial scores so that each partial score gained by a team can be related to specific functionalities. The details are not shown here for lack of space.

In Table 1 we report the percentage of the available scores for each functionality gained by the teams since 2008. The first value is the maximum score achieved by a team and the second value is the average score achieved by the finalist teams. Usually there are five finalists.

This table allows for many considerations, including: 1) which abilities have been most successfully implemented by the teams; 2) how difficult are the tests with respect to such abilities; 3) which tests and abilities need to be modified in order to guide future development in desired directions.

| Ability | 2008 [%] | 2009 [%] | 2010 [%] | 2011 [%] | 2012 [%] |
|---|---|---|---|---|---|
| Navigation | 40 / 25 | 47 / 40 | 33 / 20 | 61 / 26 | 52 / 23 |
| Mapping | 100 / 44 | 100 / 92 | 21 / 10 | 33 / 10 | 10 / 4 |
| Person Recognition | 32 / 15 | 69 / 37 | 57 / 23 | 48 / 16 | 62 / 15 |
| Person Tracking | 100 / 81 | 100 / 69 | 100 / 72 | 100 / 76 | 62 / 33 |
| Object Recognition | 29 / 8 | 39 / 23 | 6 / 1 | 35 / 10 | 56 / 20 |
| Object Manipulation | 3 / 1 | 48 / 23 | 29 / 8 | 49 / 21 | 73 / 27 |
| Speech Recognition | 87 / 37 | 89 / 71 | 50 / 38 | 76 / 59 | 90 / 56 |
| Gesture Recognition | 0 / 0 | 0 / 0 | 62 / 26 | 100 / 49 | 88 / 37 |
| Cognition | - | - | 17 / 3 | 68 / 24 | 32 / 8 |
| **Average** | **41 / 21** | **61.5 / 44.4** | **41.6 / 22.4** | **63.3 / 32.5** | **58.2 / 24.8** |

Table 1: Achieved scores for the desired abilities. The first value is the maximum score from a team, the second is the average score of the finalist teams.

From this table, two important aspects are evident: i) the effects of substantially changing the tests every two years is reflected in the scores: a general increase of performance in all the functionalities from 2008 to 2009 and from 2010 to 2011, due to the fact that tests were similar in these pairs of years, while in 2010 and 2012 tests changed in many aspects making them more difficult; ii) some functionalities have almost been completely solved while others remain difficult.

In fact, teams obtained good results in navigation, mapping, person tracking and speech recognition (average above 50%, except for navigation). Notice that the reason for a low percentage score in navigation is not related to inabilities of the teams, but to the fact that part of the navigation score is available only after some other task was achieved. The good results for speech recognition is very relevant since the competition environment is much more challenging than a typical service or domestic application due to a large amount of people and a lot of background noise. On the other hand, the results for mapping and person tracking are due to the fact that they were not applied in a very challenging setup. Person recognition and tracking performance is acceptable and thus in 2010 we increased the difficulty in this functionality during the tests. For example, more unknown people are present during the tests and a person is passing between the robot and the guide during person following. This is to test the robustness of the developed methods. Object manipulation had the highest increase of performance in 2009. Although more robust solutions have been developed by the teams, there is still some work to be done. Therefore, for 2010 we did not increase the difficulties of object manipulation in the tests. Object recognition is also reaching good performance and will not become more difficult. Until 2009 gesture recognition was not implemented by teams and the increase of available score was not sufficient to motivate teams to work in this direction. Thus for 2010, we created situations where speech is likely to fail and gesture was the only practical solution to solve the problem, obtaining the result that teams had to solve this problem and integrate this feature into their systems as well. Finally, cognitive abilities have been introduced since 2010 and good progresses have been already demonstrated.

Summarizing, by analyzing the results of team performance, it is possible to decide about the future development of the benchmarks. Possible adjustments are: 1) increasing the difficulty if the average performance is high; 2) merging of abilities into high-level skills, and more realistic tasks; 3) keeping or decreasing difficulty if the observed performance is not satisfying; 4) introducing new abilities and tests.

As the integration of abilities will play an increasingly important role for future general purpose home robots, this aspect should be especially considered in future competitions.

Other important parameters to assess the success of a benchmark are the number of participating research groups (teams) and the general increase of performance over the years. Obviously, it is difficult to determine such measures in a quantitative way: the constant evolution of the competition with its iterative modification of the rules and of the scores do not allow a direct comparison. However, it is possible to define some metrics of general increase of performance. They are based on the capability of a team to gain score in multiple tests, thus showing the effectiveness not only in implementing the single functionalities but also in integrating them into a working system, which includes the realization of a flexible and modular architecture allowing for the execution of different tasks.

In Table 2, the number of teams participating in the international competition is shown in the first row. The league has received a lot of interest since the beginning (2006). Wee registered a general increase, up to 24 teams (out of 32 requests) participating to RoboCup@Home 2010. The number of participating teams is influenced by external factors such as on which continent the benchmarks are performed and the economic situation of a team. The second row shows the increase in the total number of tests executed by all the teams during the competition (not all teams participate to all tests). The execution of over 100 tests since 2009 and some of them outside the competition arena (i.e., in a real environment) since 2010 confirms the significance of the statistical analysis we are performing. The third row contains the percentage of successful tests, i.e., tests where some score greater than zero was achieved, showing a significant and constant increase in the years compared also to the general increase of the difficulty of the tests. Finally, the fourth row contains the average number of successful tests for each team. This is a very important measure, since the enormous increase from 1.0 tests in 2006 to 7.3 in 2009 (and 6.3 in 2010 with more difficult tests and 4.2 in 2012 with even more difficult ones) is a strong indication for an

| Measure | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|
| Number of teams | 12 | 11 | 14 | 18 | 24 | 19 | 18 |
| Total amount of tests | 66 | 76 | 86 | 127 | 164 | 141 | 108 |
| Percentage of succ. tests | 17% | 36% | 59% | 83% | 74% | 73% | 58% |
| Avg. succ. tests p. team | 1.0 | 2.5 | 4.9 | 7.3 | 6.3 | 6.5 | 4.2 |

Table 2: Measures indicating general increase of performance.

average increase in robot abilities and in overall system integration. A team successfully participating in an average of 7 tests (that are quite different to each other) demonstrates not only effective solutions and implementation of all the desired abilities, but also a flexible integrated system that has important features for real world applications. Notice that in this table all the teams were considered (not only the finalists).

The results obtained by the analysis reported here clearly show that our methodology of dynamic and integrated benchmarking is producing a quick and significant progress in domestic service robotics.

## General Purpose Service Robots

Benchmarking cognitive abilities of intelligent robots is a fundamental step towards their deployment but at the same time a very difficult task to undertake. The main difficulty is due to the fact that in order to test cognitive abilities of a real system many related problems must be solved (e.g., perception and locomotion). This is the reason why there are no benchmarks for cognitive abilities of real entire systems, but only benchmarks for single components (e.g., planning competitions).

RoboCup@Home provides an experimental setting that is able to overcome this difficulty. As teams have developed service robotic systems that must demonstrate the effectiveness of the basic functionalities in other tests, we can define a test focusing only on cognition. This test, called 'General Purpose Service Robots' (GPSR), has been introduced in 2010 and focuses on the following aspects.

- There is no predefined order of actions to execute. This is to slowly transition away from state machine-like behavior programming.

- There is an increased complexity in speech recognition compared to the other tests. Possible commands are less restricted in both actions and arguments. Commands can include multiple objects, e.g., "put the mug next to the cup on the kitchen table"

- The test is about how much the robot understands about the environment and aims for high-level reasoning.

The task is executed as follows. The robot enters the arena by driving to a specified location. A task is generated on-the-fly by a computer program randomizing actions, objects and locations. The resulting command may contain different synonyms for the same action. The exact command as generated by the computer program is given to the robot by an operator (not a team member). The robot then starts executing the task. Three tasks can be solved in 10 minutes and

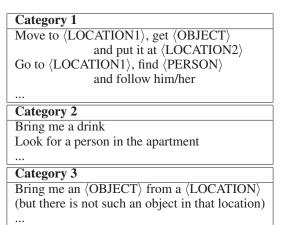| **Category 1** |
|---|
| Move to ⟨LOCATION1⟩, get ⟨OBJECT⟩ and put it at ⟨LOCATION2⟩ |
| Go to ⟨LOCATION1⟩, find ⟨PERSON⟩ and follow him/her |
| ... |
| **Category 2** |
| Bring me a drink |
| Look for a person in the apartment |
| ... |
| **Category 3** |
| Bring me an ⟨OBJECT⟩ from a ⟨LOCATION⟩ (but there is not such an object in that location) |
| ... |

Table 3: Examples of template tasks.

partial scores are available for accomplishing also portions of the task. It is important to notice that basic actions, objects and locations are known to the robot, thanks to the participation in the previous tests. However, the robot does not know in advance the exact sequence of actions to be accomplished. Also notice that the robot can communicate with the referees asking for more information, when needed, thus planning "speech" actions to acquire more knowledge on the tasks to be executed.

The requirements that all the robots are assumed to have are: a set of template tasks $T$ that can be solved by combining basic actions and functionalities already tested in previous tests during the competition (e.g., following a person, finding and recognizing people/objects, going to a location, grasping and delivering objects, speech interaction with a person, etc.), a set of people $P$ (i.e., names of persons known to the robot), a set of objects $O$ (known to the robots), and a set of locations $L$ (known to the robots).

The terms used in the task descriptions are not given beforehand to the teams, thus speech recognition and the cognition of commands need to be robust enough for the use of synonyms to describe the basic actions. For example, "go to", "move to", and "drive to" all describe a navigation command. Moreover, objects are grouped into classes (known to the robots) such as drinks, food, etc.

Each template task can have one or more locations, people, and objects as arguments. Tasks are also grouped into three categories that will be explained later. Depending on the category, commands may include a sequence of tasks, an underspecified task, or erroneous information and hence an impossible task. Some examples are provided in Table 3.

| Year | Nr. teams | Non-zero score | Avg. score |
|------|-----------|----------------|------------|
| 2010 | 7 | 2 | 3% |
| 2011 | 19 | 7 | 24% |
| 2012 | 7 | 3 | 8% |

Table 4: Evolution of results for 'GPSR' test.

A task $\bar{t}$ is the instantiation of a template task $t \in T$ obtained by selecting random locations, people and objects as needed. For example, a task generated from the first template task in Table 3 could be "Move to the kitchen shelf, get a coke and put it on the dining table".

The tasks are grouped in three categories of increasing difficulty. In the first category, the tasks are formed by a sequence of basic actions. The main difficulty for the robot is to understand a complex sequence of actions and put them together at run-time to accomplish the task. In the second category, some information is not given to the robot, for example the location in which the action must be performed is not specified, or only the object group instead of a particular object is mentioned. In this case the robot has to provide the capability of either asking for explanations to have more detailed information (e.g., "which beverage should I bring you?") or to act according to the incomplete specification (e.g., "finding and offering different beverages"). In the third category, the task contains erroneous information and it cannot be accomplished. For example, the robot is asked to find an object in a location, but there is no such object at the given location. The robot has to come up with an explanation for not having solved the task. It has to report to the operator that something went wrong, and what exactly went wrong. That is, by conducting the task the robot has to find out which part of the information in the given command has been erroneous and constitutes a conflict between the expected and actual state of the world. Also, the robot can come up with an alternative solution such as searching for the object at another location, or offering a different but similar object, e.g., another beverage. It must be noted that the robot is not aware of the command category beforehand.

Results for this test are shown in Table 4, in terms of non-zero scores and of the average percentage of score gained by teams with respect to the maximum amount available. Note that in this table we report the total test score considering both the cognition and the actual abilities required within the task. Results for 2011 take into account the fact that the test was split into two parts. Some teams have participated in both parts and thus are counted twice.

In general, it is possible to notice a slight increase of performance over the years, although the test is still very difficult. As a consequence of this difficulty, the test will not change significantly in the next years providing for chances of better overall results. The decline in performance for 2012 can be explained by the overall drop in scores due to increased difficulty of the tested capabilities.

## Robots and teams

The robots demonstrated in RoboCup@Home are as diverse as the research foci of the research groups behind the participating teams. Predominantly, teams are constituted of researchers working on projects related to human-robot-interaction, mobile manipulation, and domestic service robotics in general as well as students directly working on problems addressed in RoboCup@Home in the form of practical courses, seminars or theses.

In order to present an overview on the diversity of problems addressed in RoboCup@Home and to provide a proof-of-concept for competitions as a benchmark for integrated systems, we present various capabilities successfully demonstrated by teams in the RoboCup@Home league. It is important to notice however that, although we focus on descriptions on single functionalities, they have been demonstrated (and evaluated for performance) within an integrated system executing complex integrated tests.

### Detecting, tracking and following humans

A common approach for detecting individual persons not standing closely together is to detect (pairs of) legs in 2D laser range scans. A very robust and efficient approach to laser-based people detection has been demonstrated by Team NimbRo from the University of Bonn in Germany. They use multiple 2D laser range scanners mounted in different heights and detect and track legs and torsos (Droeschel et al. 2011). In addition, tracked hypotheses are only considered as humans once they could be verified through, respectively, vision and facial detection at the tracked position.

People detection and gesture recognition based on 3D data (e.g., by using RGB-D cameras) has also been studied. Team NimbRo uses an approach that first segments the different body parts in acquired 3D point clouds, and then estimates body (part) postures. With this ability the team's robots can perceive locations in the environment where a human operator is pointing to and recognize and grasp objects shown to the robot (Droeschel et al. 2011).

For recognizing and tracking people when following a human guide through populated areas, team RobotAssist from the University of Technology Sydney in Australia has presented an efficient yet robust approach based on 3D features (Kirchner, Alempijevic, and Virgona 2012). The team analyzes the size and shape of the people's head to shoulder region in 3D point clouds. The approach is not only very fast, but also allows for recognizing people standing with their back towards the robot. Thus the robot can constantly verify that it is following the correct person and it also does not need to request the guide to step in front of the robot (facing the robot) for re-verification after the robot has lost sight of the guide (as has been part of the *"Follow me"* in RoboCup@home until 2011.

Team B-it-bots from Bonn-Rhein-Sieg University in Germany has defined a novel 3D feature descriptor based on local surface normals and a classifier for detecting persons (Hegger et al. 2012).

Especially in populated places, tracking and following a human guide becomes intractable in situations where the

guide leaves the direct sight of the robot and the robot is not able to recover and find the user again. Team GOLEM from the Universidad Nacional Autónoma de México demonstrated successful speaker localization on a mobile robot (Rascon and Pineda 2012) during RoboCup 2012 in Mexico City. A human operator guiding the robot through an (unknown) apartment repeatedly left the robots sight and hid somewhere in the robot's vicinity. Solely by calling the robot, they could continue the guiding task.

Team eR@sers from Japan has successfully demonstrated (human) motion learning. Their approach estimates reference points, intrinsic coordinate system types and, finally, probabilistic motion model parameters for generalizing motions. For learning and generating trajectories Hidden Markov Models have been used (Sugiura et al. 2011).

## Detecting, learning and recognizing objects

Besides human-robot interaction, mobile manipulation and interacting with objects constitute a bigger part of the abilities being tested for in RoboCup@Home. Of particular relevance is fast and reliable perception of objects, and the robot's ability to safely pick up, place and hand over objects.

A very fast yet robust approach to detecting objects has been demonstrated by Team NimbRo. They approach the location where they expect the object to be, e.g., on a table, and capture the scene using a RGB-D camera. For detection and localization of objects in real-time, they use a fast method to compute local surface normals and restrict all further processing to horizontal planes (Holz et al. 2011). This method is combined with very efficient approaches to grasp planning which allows the team's robots to pick up objects only seconds after the robot has arrived at the object's approximate location (Stückler, Holz, and Behnke 2012). For modeling and tracking objects in real-time they developed fast methods using multi-resolution surfel maps. They successfully demonstrated a robot tracking a table in 3D and carrying it cooperatively with a human (Stückler, Holz, and Behnke 2012).

For on-line learning of objects (e.g., when a human user wants to teach the robot a new object), team eR@sers has successfully demonstrated the use of object names that are unknown to the robot and not contained in the robot's vocabulary. To efficiently learn out-of-vocabulary words they use phoneme sequence (Nakamura et al. 2012). For interacting with the user using the learned word, they convert the waveform to the robot's voice by EigenVoice GMM. The same approach has also been successfully used in RoboCup@Home to learn (and recognize) names of previously unknown persons.

## Conclusion

RoboCup@Home provides the environment, methodology and community effort to discuss potential benchmarks in the field of domestic service robotics. It creates dynamical problems in order to drive the intelligent robotics research agenda. By executing the benchmarks, the community has gained experience in how to set up and evaluate the benchmarks using statistical procedures. An outcome of RoboCup@Home that the authors deem important is the benchmarking of robot cognition, which requires robots that can already operate in uncertain circumstances. Many of these robust robots have been developed using the criteria set forth by the experts of the technical committees in the past seven years. Although the benchmarking of domestic service robots and of robot cognition is still a young field of research, the progress of the benchmark and the participating robots has been impressive.

## References

Droeschel, D.; Stückler, J.; Holz, D.; and Behnke, S. 2011. Towards joint attention for a domestic service robot – person awareness and gesture recognition using time-of-flight cameras. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1205–1210.

Feil-Seifer, D.; Skinner, K.; and Mataric, M. J. 2007. Benchmarks for evaluating socially assistive robotics. *Interaction Studies* 8(3):423–439.

Hegger, F.; Hochgeschwender, N.; Kraetzschmar, G. K.; and Ploeger, P. G. 2012. People detection in 3D point clouds using local surface normals. In *Proceedings of the RoboCup International Symposium (CD-ROM Proceedings)*.

Holz, D.; Holzer, S.; Rusu, R. B.; and Behnke, S. 2011. Real-time plane segmentation using rgb-d cameras. In *Proceedings of the 15th RoboCup International Symposium*, volume 7416 of *Lecture Notes in Computer Science*, 307–317. Istanbul, Turkey: Springer.

Kirchner, N.; Alempijevic, A.; and Virgona, A. 2012. Head-to-shoulder signature for person recognition. In *Proceedings of the IEEE international Conference on Robotics and Automation (ICRA)*, 1226–1231.

Nakamura, T.; Sugiura, K.; Nagai, T.; Iwahashi, N.; Toda, T.; Okada, H.; and Omori, T. 2012. Learning novel objects for extended mobile manipulation. *Journal of Intelligent and Robotic Systems* 66(1-2):187–204.

Pineda, L.; Meza, I.; Avilés, H.; Gershenson, C.; Rascón, C.; Alvarado, M.; and Salinas, L. 2011. IOCA: interaction-oriented cognitive architecture. *Research in Computer Science* 54:273–284.

Rascon, C., and Pineda, L. A. 2012. Lightweight multi-direction-of-arrival estimation on a mobile robotic platform. In *Proceedings of the International Conference on Signal Processing and Imaging Processing, held under the World Congress on Engineering and Computer Science*.

Stückler, J.; Holz, D.; and Behnke, S. 2012. Robocup@home: Demonstrating everyday manipulation skills in robocup@home. *Robotics & Automation Magazine, IEEE* 19(2):34–42.

Sugiura, K.; Iwahashi, N.; Kashioka, H.; and Nakamura, S. 2011. Learning, generation and recognition of motions by reference-point-dependent probabilistic models. *Advanced Robotics* 25(6-7):825–848.

Wisspeintner, T.; van der Zant, T.; Iocchi, L.; and Schiffer, S. 2009. RoboCup@Home: Scientific competition and benchmarking for domestic service robots. *Interaction Studies* 10(3):393–428.